<arttitle> Caveat Emptor: The Meaning of Perception and Integration in Speech Perception

<aug> Dominic W. Massaro

<aff> Department of Psychology, University of California, Santa Cruz, Santa Cruz, CA 95064.

<footnote> Present address: Department of Psychology, University of California, Santa Cruz, Santa Cruz, CA 95064.

<abs> A recent letter[1] claimed integration of auditory and tactile information in speech perception. Although I have been an advocate of multisensory integration, neither perception nor integration was sufficiently formalized, operationalized, and tested to support this claim.

<p> In a recent letter by Gick and Derrick[1], called *Aero-Tactile Integration in Speech Perception*, the authors concluded that "perceivers integrate naturalistic tactile information during auditory speech perception without previous training." This conclusion was made based on their findings that inaudible air puffs on the skin increased the accuracy of distinguishing between speech sounds that differed in aspiration (p versus b and t versus d). Notwithstanding the fact that the integration of audible and visible information has been repeatedly and conclusively demonstrated[2], this note questions the authors' conclusion that "These results demonstrate that perceivers integrate event-relevant tactile information in auditory perception in much the same way as they do visual information."

<p> This critique is based on two important aspects of behavioral science that are often assumed without question but, on deeper analyses, should be questioned. First, investigators often make the assumption that participants in an experiment perform the task exactly as they were instructed. In the Gick and Derrick experiment, "Participants were … asked to identify by pressing a button whether they heard 'pa' or 'ba' in the labial block, and 'ta' or 'da' in the alveolar block." In this case, the investigators are making the questionable assumption that the participants' identification response was

based on their auditory experience of the speech sound rather than on a cognitive decision based on what they know about speech and the task. Participants could have easily viewed the task as a problem-solving task in which they applied some other strategy than basing their decision on their auditory experience. One test of this possibility would be to present only the puff of air without the speech sound with the same instructions. If participants are more likely to identify the puff with the aspirated alternative, then we learn that differences in identification judgments do not necessarily mirror only what the participants heard. There are several more manipulations that have been somewhat successful in pinpointing perceptual experience in the study of auditory-visual speech but it safe to say that experience remains perhaps an impenetrable barrier in behavioural science[3,4].

<p> A second caveat is that the term “integration” tends to be used very loosely in cognitive and neuroscience, even though one aspiring goal of science is to be very precise in its terminology. Dictionary definitions are naturally very broad to account for the fundamental fuzziness of language use. For me, integration might mean “the act of combining into an integral whole,” “formed into a whole or introduced into another entity” or “become one”[5]. In behavioral science, however, a term should be formalized within a precise and testable model. In our research[4,6], we make an important distinction between integration and non-integration processes. Multi-sensory integration involves combining continuous information from two sensory modalities so that the perceptual experience to a given multi-sensory event reflects the contribution of both modalities. In auditory-visual speech perception, for example, the pairing of two different syllables often produces a unique identification. An auditory ba paired with a visual da often produces va or tha identifications; and an auditory da paired with a visual ba often produces bda identifications, whereas the opposite pairing never produces dba responses[4,6]. A non-integration process involves an identification of a multi-sensory event that is based on just one of the sensory inputs, which could explain the effect of air puffs on subjects' decisions. Distinguishing between these alternatives is not easy and usually requires more elaborate experimentation and model testing than researchers typically carry out[7].

<p> The improved performance in the Gick and Derrick experiment with the air puff could easily have occurred without an underlying integration process. One outcome that

supports this interpretation is that the location at which the puff was delivered did not matter. The authors had rationalized that a puff on the hand would be more relevant to the perceivers' experience with speech than a puff on the neck. A non-integration process seems more likely, given that puffs at both locations improved performance. A second reason to be sceptical of the conclusion that integration occurred is that the effectiveness of tactile aids for supplementing hearing loss requires a significant amount of learning[8].

<p> To conclude, using the strict criteria described here, there is indeed evidence that perceivers integrate auditory and visual information from the face in speech perception[2,4,6]. The analogous conclusion for auditory and tactile information remains to be demonstrated.

<bibcit> 1. Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. Nature, 462. 502-504.

<bibcit> 2. Massaro, D.W. (2004) From Multisensory Integration to Talking Heads and Language Learning. In G. Calvert, C. Spence & B. E. Stein (Eds.), Handbook of Multisensory Processes (pp.153-176). Massachusetts: MIT Press.

<bibcit> 3. Ariely, D. (2008). Predictably Irrational: The Hidden Forces That Shape Our Decisions. HarperCollins.

<bibcit> 4. Massaro, D. W. (1998). Perceiving talking faces: From speech perception to a behavioral principle. Cambridge, Massachusetts: MIT Press.

<bibcit> 5. http://www.google.com/search?hl=en&client=firefox-a&rls=org.mozilla:en-US:official&hs=4wL&defl=en&q=define:integration&ei=w9IPS7WKHoyCMuXqyDM&sa=X&oi=glossary_definition&ct=title&ved=0CAcQkAE

<bibcit> 6. Massaro, D. W. (1987). Speech perception by ear and eye: A Paradigm for psychological inquiry. Hillsdale, NJ: Erlbaum.

<bibcit> 7. Massaro, D. W., & Chen, T. H. (2008). The motor theory of speech perception revisited. Psychonomic Bulletin & Review, 15, 453-457.

<bibcit> Bernstein, L. E., Demorest, M. E., Coulter, D. C. & O' Connell, M. P. Lipreading sentences with vibrotactile vocoders: performance of normal-hearing and hearing-impaired subjects. J. Acoust. Soc. Am. 90, 2971–2984 (1991).

<ack> The author thanks Greg Bryant for drawing his attention to the article and for his helpful comments on a draft of this letter.

<corr> Correspondence and requests for materials should be addressed to the author. (e-mail: massaro@ucsc.edu).